Plasma Metabolomics to Identify and Stratify Patients with Impaired Glucose Tolerance

Charlotte Wildberg, Annette Masuch, Kathrin Budde, Gabi Kastenmüller, Anna Artati, Wolfgang Rathmann, Jerzy Adamski, Thomas Kocher, Henry Völzke, Matthias Nauck, Nele Friedrich, Maik Pietzner

The Journal of Clinical Endocrinology & Metabolism
Endocrine Society

Submitted: May 13, 2019
Accepted: August 01, 2019
First Online: August 07, 2019

OGTT-prediction by plasma metabolomics

Plasma Metabolomics to Identify and Stratify Patients with Impaired Glucose Tolerance

Charlotte Wildberg[1], Annette Masuch[1], Kathrin Budde[1,2], Gabi Kastenmüller[3], Anna Artati[4], Wolfgang Rathmann[5], Jerzy Adamski[4,6,7], Thomas Kocher[8], Henry Völzke[2,9,10], Matthias Nauck[1,2], Nele Friedrich[1,2], Maik Pietzner[1,2]

[1]*Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany*

[2]*German Center for Cardiovascular Research (DZHK e.V.), partner site Greifswald, Germany*

[3]*Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany*

[4]*Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany*

[5]*Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany*

[6]*Lehrstuhl für Experimentelle Genetik, Technische Universität München, Freising-Weihenstephan, Germany*

[7]*DZD (German Center for Diabetes Research), München-Neuherberg, Germany*

[8]*Unit of Periodontology, Department of Restorative Dentistry, Periodontology, Endodontology, and Pediatric and Preventive Dentistry, Dental School, University Medicine Greifswald, Greifswald, Germany*

[9]*Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany*

[10]*DZD (German Center for Diabetes Research), site Greifswald, Greifswald, 17475, Germany*

ORCiD numbers:

0000-0003-3437-9963

Pietzner

Maik

Received 13 May 2019. Accepted 01 August 2019.

Objective Impaired glucose tolerance (IGT) is one of the pre-symptomatic states of type 2 diabetes mellitus and requires an oral glucose tolerance test (OGTT) for diagnosis. Our aims were two-fold: 1) characterize signatures of small molecules predicting the OGTT-response and 2) identify metabolic subgroups of participants with IGT.
Methods Plasma samples from 827 participants of the Study of Health in Pomerania free of diabetes were measured utilizing mass spectrometry and proton-nuclear magnetic resonance spectroscopy. Linear regression analyses were used to screen for metabolites significantly associated with the OGTT-response after two hours adjusting for baseline glucose and insulin levels, as well as important confounders. A signature predictive for IGT was established using regularized logistic regression. All IGT cases (N=159) were selected and subjected to unsupervised clustering using a k-means approach.
Results and Conclusion: In total, 99 metabolites and 22 lipoprotein measures were significantly associated with either 2-hour glucose or 2-hour insulin levels. Those comprised variations in baseline concentrations of branched-chain amino keto-acids, acylcarnitines, lysophospholipids or phosphatidylcholines largely confirming previous studies. By the use of these metabolites, IGT-subjects segregated into two distinct groups. Our IGT prediction model combining both clinical and metabolomics traits achieved an AUC of 0.84, slightly improving the prediction based on established clinical measures. The present metabolomics

approach revealed molecular signatures associated directly to the response of the OGTT and to IGT in line with previous studies. However, clustering of IGT subjects revealed distinct metabolic signatures of otherwise similar individuals pointing towards the possibility of metabolomics for patient stratification.

Comprehensive molecular profiling of plasma samples improved classification and stratification of IGT subjects over and above clinical measures.

Introduction

Type 2 diabetes mellitus is one of the major health burdens across the world [1]. It is primarily defined by the level of hyperglycaemia associated with an increased risk for microvascular and macrovascular disease development. However, type 2 diabetes mellitus has a long pre-symptomatic stage. Impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) are two states associated with insulin resistance, in which glycaemia is disturbed, but not as much to be classified as type 2 diabetes mellitus and therefore also referred to as pre-diabetes. Besides determination of HbA_{1c} and measurement of fasting plasma glucose (FPG), the OGTT is the gold standard tool to diagnose these pre-diabetic conditions [2].

Although the OGTT is a standard method to diagnose both type 2 diabetes mellitus and IGT, it has become progressively unpopular, as it is costly in terms of time and money [3, 4]. Also due to its complexity, insufficient reproducibility and overall inconvenience its use in clinical practice has declined in the past years [5], making a replacement of the OGTT desirable. More importantly, patients' HbA_{1c} and FPG may be within normal limits while glucose tolerance is not, which remains unrecognized without OGTT. However, the OGTT is currently the only method to detect isolated IGT and determination of plasma HbA_{1c} and FPG are likely to miss patients with isolated IGT [3, 4] and hence those who would benefit from early intervention strategies.

To improve patient care in terms of convenience and diagnostic safety for IGT, metabolomics studies are a promising tool. A number of studies have revealed diverse small molecules (metabolites) related to type 2 diabetes mellitus or pre-diabetes [4, 6-12]. Elevated concentrations of in particular branched-chain amino acids (BCAAs) in blood plasma were found to be predictive for incident type 2 diabetes mellitus up to ten years before disease onset [7, 8, 11, 13, 14]. Those findings were extended to further downstream metabolites, e.g. 3-methyl-2-oxovalerate or ketone bodies [15, 16]. In contrast, inverse associations, e.g. with the amino acid glycine, have been observed with respect to IGT [4, 8] and the risk of developing type 2 diabetes mellitus [6]. Besides glucose other carbohydrates like lactate, mannose, malate and arabinose were reported to be linked to insulin resistance [9, 16]. Even lipid species like lysophosphatidylcholine (18:2) and linoleoylglycerophosphocholine [4, 8], vitamins [4, 6], and other individual molecules such as acetylcarnitine and several yet to be identified metabolites were shown to be associated with IGT and/or insulin resistance [4, 6, 8, 17]. Some metabolites are most likely the result of already existing high insulin levels (e.g. glycine) while others may be related to regulatory effects in IGT-affected individuals (e.g. lysophosphatidylcholine) [8].

Numerous studies have highlighted the great potential of metabolomic approaches to improve our understanding of biochemical pathways disturbed years before the clinical manifestation of type 2 diabetes mellitus. We aimed to profile a molecular signature predictive for the OGTT response in plasma among about 800 non-diabetic subjects from the general population. Through integration of diverse metabolomics techniques, targeted and untargeted, we further investigated the presence of metabolic subgroups of IGT patients. Hence, we pay tribute to the multifaceted origin of impaired glucose homeostasis as has been

shown by a number of recent clustering studies [18-20] as well as outlined by the palette model emphasizing the different shades of type 2 diabetes mellitus aetiology [21].

Methods

Study Population

The Study of Health in Pomerania (SHIP-TREND) is a population-based study conducted in West Pomerania, a rural region in north-east Germany and a detailed description of the sampling procedure and the study population can be found elsewhere [22]. In total, 4420 subjects chose to participate (50.1% response). All participants gave written informed consent before taking part in the study. The study was approved by the ethics committee of the University of Greifswald and conformed to the principles of the declaration of Helsinki. SHIP data are publicly available for scientific and quality control purposes by application at www.community-medicine.de.

For the first 1000 subjects without self-reported type 1 or 2 diabetes, plasma metabolome data based on MS and proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy were obtained. Exclusion criteria applied included (overlap exists, Fig. 1): 1) missing values in OGTT measures or confounding variables (n=54) 2) fasting time less than eight hours (n=71) 3) newly diagnosed diabetic subjects (n=26; HbA_{1c} > 6.5 or fasting glucose > 11.1 mmol/L or intake of anti-hyperglycemic medication) and 4) exclusion after quality control of metabolomics data (n=22). Finally, a total sample of 827 subjects was included in the analyses. Despite current guidelines are clearly defining 2-hour glucose measurements of ≥ 11.1. mmol/L as diabetic conditions [2], we decided to include these subjects (n=23) in the analyses as they were identified by OGTT only and the aim of the study was to search for spot metabolic markers predicting the dynamic response to a glucose challenge. Figure 1 summarizes sample compilation and statistical analyses.

Standard Laboratory Assays

Fasting blood samples (≥ 8 hours) were collected between 6:00 am and 1:00 pm from the cubital vein of subjects in the supine position and analysed immediately or stored at -80°C in the Integrated Research Biobank (Liconic, Liechtenstein) at the University Medicine Greifswald. Directly thereafter, non-diabetic participants were given a standardized solution of glucose containing 75mg (Dextro OGT, Boehringer Mannheim, Mannheim, Germany). Serum cystatin C, lipids (total cholesterol, HDL and LDL cholesterol, triglycerides (TG)), high-sensitivity C-reactive protein (hsCRP) and catalytic activity concentration of alanine aminotransferase (ALT) were measured by standard methods (Dimension VISTA, Siemens Healthcare Diagnostics, Eschborn, Germany). Plasma insulin levels were measured (Centaur XP by Siemens Healthcare Diagnostics) and the homeostatic model assessment of insulin resistance (HOMA-IR) index was calculated as insulin (μU/mL) × glucose (mmol/L)/22.5 [23]. HbA_{1c} was determined by high-performance liquid chromatography (Bio-Rad, Munich, Germany).

Metabolome Analyses

A detailed description of all applied measurement techniques has been published before [24]. Three different approaches were combined: 1) non-targeted MS-based profiling of plasma samples as reported previously 2) targeted MS-based profiling of plasma samples using the AbsoluteIDQ p180 Kit (BIOCRATES LifeSciences AG, Innsbruck, Austria) and 3) ^{1}H-NMR-based profiling of plasma samples to derive measures of lipoprotein particles.

Quality control

Differences in metabolite measurements due to day-to-day variation in both MS-based techniques were accounted for by median-normalization for each metabolite, i.e. the median of each metabolite on each runday was used to rescale the respective metabolite

measurements yielding the same median at each runday. To this end samples have been assigned to specific runday at random prior measurements. Normalized metabolite measurements were subsequently log2-transformed and each data set was submitted to outlier identification by robust principle component analysis to exclude samples strongly deviating in their global profile. Only metabolites with less than 20% missing values were used to this end while imputing missing values with a low value. As a result 13, 4, and 6 samples have been excluded for non-targeted MS, targeted MS, and ^{1}H-NMR, respectively. Quality control for ^{1}H-NMR-derived measures of lipoproteins was further done in relation to laboratory measurements of blood lipids yielding excellent correlation coefficients (r>0.90).

Data integration

Most of the metabolites were unique to one of the applied techniques. However, 44 plasma metabolites were overlapping with both techniques. Following the grouping of metabolites in biochemical classes (i.e. lipids, amino acids and carbohydrates), correlations of those metabolites measured on both platforms were computed with all members of the same biochemical class. Subsequently, the metabolite with the higher median correlation across all class members was kept for further analysis.

After quality control and pre-processing 613 plasma metabolites were available for statistical analyses. Note that some of these could not be unambiguously assigned to a chemical identity and are referred to hereafter with the notation “X” followed by a unique number. Data on lipoprotein particles comprise 117 measures describing the gradient from VLDL particles to HDL particles, including their triglycerides, cholesterol, free cholesterol, phospholipid as well as apolipoprotein B, A1 and A2 content.

Statistical Analysis

Linear regression models were performed to assess the associations of plasma metabolites (independent variables) with 2-hour measures of plasma glucose and insulin (dependent variables). To fulfil requirements of linear regression metabolite levels and plasma insulin were log-transformed. All models were adjusted for baseline levels of glucose or insulin, age, sex, waist circumference, physical activity, smoking behaviour, serum ALT, eGFR and hsCRP. To screen for metabolites specifically associated with the presence of IGT we run logistic regression models with the same adjustment set for each metabolite as exposure and a binary IGT-variable as outcome. To account for multiple testing, the p-values from regression analyses were adjusted by controlling the false discovery rate (FDR) at 5% using the Benjamini-Hochberg procedure.

To search for putative metabotypes related to IGT, all cases (n=159) were selected and subjected to unsupervised clustering using a k-means and a hierarchical clustering analyses (HCA) approach. To this end only significantly associated metabolites in either linear or logistic regression with less than 20% missing values were evaluated. Missing values were first imputed using a k-nearest neighbour approach. to determine the optimal number of clusters, 30 different measures of cluster separation were evaluated as implemented in the R package *NbClust* [25]. To identify metabolites responsible for subdivision into the two clusters we created a binary variable identifying cluster belonging among IGT-subjects and run random forest analysis to obtain measures of variable importance.

A signature predictive for IGT (2-hour glucose > 7.8 mmol/L) among the whole study population using least absolute shrinkage and selection operator (LASSO) for variable selection was compiled. Using a two-staged cross-validation procedure allowed us to test for robustness of selected features across random subsets of the population, as well as to assess generalizability of the results following previous work [24]. Each variable was scored by average area under the curve in the final classification loop in case the variable was included in the final model. Three types of variable set ups were used to perform this classification.

Considering only clinical variables as presented in Table 1, considering metabolites only, and a combination of both. Finally, three sparse logistic regression models were built to predict IGT. Performances of the individual models were compared based on receiver operating characteristics (ROC) statistics, i.e. comparing the area under the curve (AUC) using a Delong test, and by computing the continuous net reclassification improvement (cNRI), which is a combined measure of the amount of samples correctly reassigned to either cases or controls between a standard and an updated model. Only metabolites with less than 20% missing values were included (n=432) in these analyses and imputation of remaining missing values was done using k-nearest neighbour imputation.

We calculated a data-driven metabolic network based on Gaussian graphical models using the imputed data set from prediction analyses due to their ability to recreate a sparse, biochemical mirroring, network[26].

Statistical analyses were done using R 3.3.2 (R Foundation for statistical computing, Vienna, Austria).

Results

Study population

Table 1 summarizes characteristics of the study population stratified by glucose tolerance being either normal (NGT, < 7.8 mmol/L) or impaired (IGT, >7.8 mmol/L). IGT subjects were characterized by significantly higher measures of glucose homeostasis, including HbA_{1c} and HOMA-IR as well as OGTT results for plasma insulin and glucose. Waist circumference and plasma concentrations of TG were higher in IGT subjects, whereas LDL- and total cholesterol levels did not differ considerably. A significantly higher proportion of current smokers were found in the NGT population compared to IGT subjects.

Associated metabolites

In total, plasma concentrations of 99 metabolites showed significant positive or inverse associations with 2-hour glucose and/or insulin levels (Fig. 2 and Tab. 2). The majority of metabolites were associated with 2-hour glucose levels. In detail, the amino acids glycine, betaine and asparagine were inversely associated whereas degradation intermediates of branched-chain amino acids (BCAAs), e.g. 3-methyl-2-oxovalerate or alpha-hydroxyisovalerate, were positively associated with 2-hour glucose concentrations. Members of the urea cycle, e.g. citrulline and ornithine, as well as the indoles kynurenine and indoleacetate were inversely associated. The glucose metabolites lactate and pyruvate were positively associated. Apart from carnitine itself (inversely associated) all plasma acylcarnitine species were positively associated with 2-hour glucose levels. Significant positive association with NEFAs, comprised saturated fatty acids (SFAs), such as palmitate (16:0), and 2-hour glucose/-insulin. Moreover, prominent (poly)unsaturated fatty acids (ω-3/-6 fatty acids) such as arachidonate and linoleate associated positively to both, glucose and insulin. In contrast, significant inverse associations were seen for lysolipids (1-linoleoylglcerophosphoethanolamine) and several phosphatidylcholines (PCs) with 2-hour glucose levels. Plasma levels of two lysophosphatidylcholines (lysoPC) additionally showed significant inverse associations to 2-hour insulin. A positive association was found between 2-hour glucose and peptide fragments of complement component 3 (HWESASLLR and HWESASXX*) [27], whereas significant inverse associations were observed with respect to xenobiotics including plasma metabolites associated with coffee consumption (e.g. 4-vinylphenol-sulfate, catechol sulfate, and trigonelline [28]).

Adjustment for body mass index instead of waist circumference in linear regression analysis had virtually no effect on effect estimates or p-values (Pearson correlation coefficient: 0.99).

Further, logistic regression analysis for IGT was well in line with results from linear regression analysis (Fig. 3). Metabolites unique to IGT-analysis comprised inverse associations, i.e. lower odds for IGT, with asymmetric dimethylarginine and stachydrine.

Lipoprotein subfractions

In total, 22 measures of lipoproteins were associated with 2-hour glucose levels. Inverse associations were seen with the cholesterol content of LDL particles, whereas the TG content in large VLDL and small LDL particles was positively associated. Even measures of small VLDL particles were positively associated (Fig. 3).

Clustering of the IGT population

Within the IGT population, k-means (n=88 and 71) and HCA (n=90 and 69) each revealed two clusters with robust agreement between both as indicated from Jaccard indices of 0.73 and 0.67, respectively (Fig. 4). Based on the assignment of the k-means approach participants differed slightly in their 2-hour glucose concentrations as well as in waist circumference. Thereby the latter was likely due to a higher number of males in cluster 2 (Fig. 4). All other parameters for glucose homeostasis did not differ significantly between the two clusters (Fig. 4). Random forest analysis revealed a high importance of (un-)saturated fatty acid species for the distinction between both groups, which was also evident when drawing standardized metabolite levels as a heatmap across all IGT-subjects (Fig. 4)

Predicting IGT

Three types of models were established being predictive for IGT. The classification scheme consisted of either clinical or metabolic predictors, as well as a combination of clinical and metabolic predictors for IGT. The final subset of variables evolved during variable selection is presented in Table 3. Models relying on either established clinical variables or metabolites showed comparative performance in the ROC-analysis with an AUC of ~0.79 whereas the combination of both data sets indicated a significant improvement compared to the only clinical- ($p<0.002$) or metabolite-based ($p<0.0001$) model with a final AUC of 0.84 (95%-CI: 0.81; 0.87; Fig. 5). An improvement between the purely clinical and the combined model was also obvious from evaluating the cNRI: 0.47 (95%-CI: 0.30 - 0.64), $p<0.001$. Metabolites consistently included in both type of models included alanine, lysoPC C18:2, and 2-oleoylglycerophosphocholine. Inclusion of 3-methyl-2-oxobutyrate and trigonelline further contributed to better discriminative performance. The unknown metabolite X – 11727 might be of interest from an aetiological perspective, since it established a link in the data-driven metabolic network between the sum of hexoses and urea (Fig. 5B).

Discussion

The present study investigated baseline levels of plasma metabolites associated with 2-hour glucose and/or 2-hour insulin following an OGTT and identified 99 small molecules to be associated independent from steady state glucose homeostasis. We replicated a number of already known metabolites related to (pre-)diabetes such as BCAAs-derivatives, acylcarnitines, fatty acids or measures of lipoprotein particles, but also observed novel small molecules, e.g., a fragments of complement component 3 (C3), several unknown compounds or heme degradation products. Based on these metabolomics markers the subjects with OGTT-proven IGT segregated into two distinct clusters with only minor differences in 2-hour glucose concentrations. The clustering was in particular due to strong differences in plasma NEFAs. Compared to readily available clinical measures, incorporation of five metabolites slightly but significantly improved classification of IGT-subjects using baseline measures only.

Replication of previously described signatures

Elevated plasma levels of BCAAs have been repeatedly reported to be associated with either incident diabetes or pre-diabetic states [4, 7, 8, 11-14, 16, 29, 30] and our results with respect to BCAA-derivatives, such as 3-methyl-2-oxovalerate, largely confirm these observations. We like to note, that BCAAs themselves did not contribute to our IGT prediction models which aligns with previous reports [8, 30]. Briefly, a lack of insulin-mediated suppression of skeletal muscle proteolysis might explain the elevated levels of BCAAs in (pre)diabetics [31]. Even the contribution of gut microbiota, synthetizing BCAA, has been suggested [32]. Both mechanisms as well as possible yet to be identified ones raise BCAA concentrations in plasma and thereby likely the concentration of their first degradation product: branched-chain alpha-keto acids (BCKAs). BCKAs in turn presumably reduce pyruvate dehydrogenase and α-ketoglutarate dehydrogenase activity [31], which are essential in the tricarboxylic acid cycle [33]. Subsequently, pyruvate metabolism shifts towards lactate and alanine formation, which aligns with our observation of elevated baseline levels of these two metabolites. Notably, alanine served as a significant variable in the metabolite-based and combined IGT prediction model. Alanine is strongly linked to glucose- and BCAA-metabolism [33, 34]. In the state of insulin resistance alanine is deliberated by the skeletal muscle cells to maintain glucose homeostasis, as it is then converted into pyruvate, a source for augmented gluconeogenesis in the liver [33]. The lack of association between primary BCAAs and 2-hour glucose levels might be due to the mediating effect of hepatic fat accumulation (using serum ALT as surrogate marker in linear regression analysis) as has been shown previously for the same subjects [24]. However, genetic evidence [35, 36] implies BCAA metabolism to be on a causal path to type 2 diabetes *via* insulin resistance and experimental evidence suggested BCAA-mediated impaired signalling through mechanistic target of rapamycin (mTOR) uncoupling downstream signal transduction of insulin as one possible mechanism [31].

Consistent with previous studies [4, 8, 37, 38] baseline levels of lysoPCs (18:1 and 18:2) were inversely associated with both 2-hour glucose and -insulin concentrations in the present study. Predominantly the lysoPCs 18:2, 18:1 and 18:0 have been reported in diabetes related studies, but chiefly lysoPC 18:2, also referred to as linoleoyl-glycerophosphocholine (L-GPC). Briefly, lysoPCs are catabolites of phosphatidycholines (PCs) present in cell membranes or on the surface of lipoprotein particles. PC degradation is catalysed by phospholipase A2 (PLA2) [8]. Notably, lower mRNA levels of cytoplasmic PLA2 have been observed in whole blood samples of IGT- and type 2 diabetes mellitus-affected subjects [8], which might account for the lower concentrations of lysoPCs in (pre-)diabetic patients.

A number of ether and ester-PCs were significantly inversely associated with 2-hour glucose, with the exception of PC aa C36:4 showing a positive association partially replication a previous observation among older adults [39]. PCs are essential for the assembly and secretion of lipoproteins and even regulate the amount of lipoproteins being released into circulation [40]. Hence, they were closely linked to the lipoprotein profiling applied in the present study. Insulin resistance is known to have strong effects on lipoprotein metabolism including TG-enrichment of small LDL-particles, likely due to an increased hepatic secretion of TG-rich VLDL particles. A process tightly linked to a decline in the cholesterol content of HDL-particles [41, 42]. The latter, however, was not obvious from the present analyses and we observed an increased TG content of HDL3-particles instead. An increase in cholesterol ester transfer protein (CETP) activity – which mediates the transfer of cholesterol esters from HDLs to TG-rich lipoproteins in exchange for TG [43] – has been suggested in IGT subjects [42, 44], and may account for the positive association with the TG-content in HDL3 particles. Another cross-link between lipoprotein determination and small molecules might have accounted for the positive association between ω-3 and ω-6 polyunsaturated fatty acids (PUFAs) and 2-hour glucose and insulin measures. Briefly, the essential ω-3 PUFAs cannot

be synthesised by mammals and hence their content in the human body entirely relies on exogenous intake, e.g. from fish or plant oil intake. Previous studies already emphasized a predictive value over and above the biochemical determination of such food bio makers compared to food questionnaires with relation to type 2 diabetes risk prediction [45]. However, cross-linking small molecule profiling with high-resolution lipoprotein profiling using ^{1}H-NMR allows us to speculate about an additional explanation, namely the increased secretion of TG-rich VLDL particles which was strongly positively correlated with plasma PUFA levels [24]. Hence, hepatic co-secretion of PUFAs as part of TG-rich VLDL particles in states of peripheral insulin resistance might be another explanation for this finding.

Novel metabolites pointing towards inflammation

Apart from the previously reported metabolites related to insulin resistance, pre-diabetic states and type 2 diabetes mellitus, the present study revealed a few novel species. We observed that L-urobilin and bilirubin, degradation products of haemoglobin, were positively associated with 2-hour glucose, but not with 2-hour insulin. Bilirubin has previously been characterized to be anti-oxidative and anti-inflammatory and therefore being a protective factor with respect to peripheral vascular diseases [46, 47]. The heme oxygenase system, which is essential for haemoglobin metabolism might be overactive in (pre-)diabetic individuals, as it possibly augments insulin sensitivity and glucose uptake [47, 48]. In addition, we observed a positive association to 2-hour glucose for fragments of the complement component 3 (C3). Notably, a longitudinal study previously found plasma C3 levels to be positively associated with insulin resistance, 2-hour glucose and fasting plasma glucose and even incident type 2 diabetes mellitus [49]. Activation of the innate immune system aligns with other pro-inflammatory effects such as toll-like-receptor activation by fatty acids [50, 51] and high CRP levels [52] associated with insulin resistance and type 2 diabetes mellitus, respectively [53]. These findings are somewhat remarkable, since all regression models were adjusted for hsCRP levels, which might indicate that proteins like C3 indicate alternative paths how the inflammatory response interferes with insulin sensitivity.

Clustering of IGT subjects

Within the group of participants with IGT, two clusters of subjects were identified using k-means clustering and partially verified by HCA, which differed by a distinctive fatty acid signature. One cluster was characterized by elevated baseline levels of a number of (un-)saturated fatty acids being long- or medium-chained and additionally several acylcarnitines. The clusters did not differ in fasting plasma glucose, HbA_{1c} or insulin resistance (HOMA-IR). However, 2-hour glucose concentrations were slightly different and hence we cannot completely rule out that members of the cluster with higher baseline plasma concentrations of fatty acids have already reached an advanced stage of IGT. Notably, the difference in fatty acid levels clearly exceeded the difference in 2-hour glucose levels between the clusters (Fig. 4) but participants with likely diabetic 2-hour glucose concentrations (>11.1 mmol/l) were significantly enriched in cluster 1 (Fisher's exact test, 3.2-fold, p=0.02). Recent studies already identified subgroups in type 2 diabetes mellitus-affected individuals [18-20]. These clusters differed in disease progression [18] and were characterized by distinct diabetic complications such as nephropathy, retinopathy and cardiovascular diseases [18, 19]. Another genetic approach, using single nucleotide variants previously associated with an increased risk of type 2 diabetes mellitus, found clusters of those variants divergent in beta cell function, their features of insulin resistance (BMI, waist circumference, lipoprotein and TG profile) and liver lipid metabolism [20]. These findings lead to the assumption that subjects of the presented IGT clusters may – when developing diabetes – diverge into distinctive type 2 diabetes mellitus subgroups.

The potential clinical value of this finding, however, could only be verified using longitudinal data on important outcome characteristics, including time-of-onset of type 2 diabetes mellitus and related complications, responsiveness to treatments schemes or more importantly effectiveness of prevention schemes including changes in nutrition and lifestyle. Successful relation of such longitudinal outcomes for specific IGT subgroups might indicate the usefulness of metabolomics characterization of patients at risk for type 2 diabetes mellitus in a tailored fashion.

With respect to a possible mechanistic interpretation it has previously been suggested that an incomplete or impaired β-oxidation accompanied by accumulation of NEFAs and acylcarnitines is associated with insulin resistance and IGT [54]. Particularly long-chain fatty acids such as palmitic acid seem to be a putative mediator [54, 55]. SFAs, such as palmitic acid, have been described to activate toll-like receptors leading to inhibition of insulin signal transduction among others [50, 51]. Moreover, the stimulation of toll-like receptors is related to the activation of the innate immune system. It therefore links to hsCRP as a positive IGT predictor and to higher levels of complement component fragments in plasma specimen of IGT affected subjects, in line with our observations. However, even insulin resistance could be a hidden driver, since a genetic prediction study has suggested a causal effect of insulin resistance on plasma levels of palmitoleate and oleate [56].

IGT Prediction

Applying extensive feature selection revealed three IGT prediction models based on different clinical and metabolic features. While the two models based on either metabolites or clinical traits performed with an approximately equivalent accuracy, our combined feature selection approach using metabolome and clinical traits improved prediction of IGT.

Regarding the clinical traits in our combined model, established diagnostic parameters, including fasting plasma glucose, HOMA-IR and advanced age as a risk factor were of predictive value whereas HbA_{1c} as a measure long-term exposure to high glucose concentrations was not selected. By variable selection, metabolites already discussed in previous sections, which have been described as IGT-specific such as lysoPC 18:2 (L-GPC) and the BCKA 3-methyl-2-oxobuytrate were included [4, 8, 30]. The strong value of lysoPC 18:2 might be in close relation to the frequent observation of the influence of a low-grade inflammatory state as indicated by higher CRP values on the OGTT outcome [57]. However, as already outlined above, the precise interplay between an inflammatory environment, PLA(2) activity and insulin signalling remains to be established. High plasma concentrations of alanine have been previously described to precede adverse OGTT outcomes by up to 6.5 years [13] but are less informative for prediction of type 2 diabetes onset [14]. A strong relation of plasma concentrations of alanine with glucose homeostasis is likely explainable as alanine can be utilized as an alternative source for glucose during gluconeogenesis. States of high (intracellular) glucose availability or suppressed gluconeogenesis, for instance in a still insulin-sensitive liver despite peripheral insulin resistance, may induce less uptake of alanine in peripheral tissues and hence higher concentrations in the circulation. Further, net production might be increased as well, as indicated by higher levels of pyruvate in relation of 2-hour glucose concentrations. In general, the prediction procedure revealed the metabolic heterogeneity leading to an impaired response in an OGTT, presumably caused by either impaired insulin secretion, i.e. beta-cell function or insulin resistance.

Strengths and Limitations

The strength of the present study lies in the non-targeted metabolomics approach regarding various metabolic species including novel metabolites and also taking lipoproteins into account. Even though we achieved an improvement in the prediction of IGT using metabolites, this needs further validation. We have to note that the rather small number of

IGT subjects in the present study did not permit us to investigate a further distinction from IFG, which might lead to further stratification of patients. In general, replication of our results and clinically usefulness of the clustering approach with respect to time-of-onset of type 2 diabetes, treatment outcomes or progression of associated co-morbidities specific to each cluster have to be validated in independent cohorts. Same holds true for the evaluation of the stability over time in candidate metabolites from IGT prediction.

Conclusion

The present metabolomics approach provides detailed insights on various metabolites associated with the outcome of the OGTT, with a particular focus on BCAA-catabolites, acylcarnitines, lysophospholipids, PCs and lipoproteins. Within the IGT-subject group, we identified two clusters with a distinctive fatty acid signature, possibly reflecting either differently advanced stages of IGT or distinct disease aetiologies. An IGT prediction model based on a combination of metabolic and clinical traits was established being superior to models including either metabolic or clinical traits but improvement might not be of direct clinical relevance. However, further studies are needed to validate our findings and prove the applicability of metabolic profiling to improve diagnosis of disturbed glucose metabolism.

Funding SHIP is part of the Research Network of Community Medicine of the University Medicine Greifswald, which is supported by the German Federal State of Mecklenburg- West Pomerania. This work was funded by grants from the German Federal Ministry of Education and Research (BMBF, grants 01ZZ0403, 01ZZ0103, 01GI0883, AtheroSysMed 03IS2061B), the Ministry for Education, Research and Cultural Affairs, as well as the Ministry of Social Affairs of the Federal State of Mecklenburg-West Pomerania. This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED). The GANI_MED consortium is funded by the Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg-West Pomerania (03IS2061A).

German Federal Ministry of Education and Research, 01ZZ0403, Henry Völzke;
German Federal Ministry of Education and Research , 01ZZ0103, Henry Völzke;
German Federal Ministry of Education and Research , 01GI0883, Henry Völzke;
German Federal Ministry of Education and Research , 03IS2061B, Not Applicable;
Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg-West Pomerania, 03IS2061A, Not Applicable

Address correspondence to: Maik Pietzner, University Medicine Greifswald, Institute of Clinical Chemistry and Laboratory Medicine, Ferdinand-Sauerbruch-Str., 17475 Greifswald, E-Mail: maik.pietzner@med.uni-greifswald.de, Tel: +49-3834-86-19659

Contribution statement

CW, MP and AM wrote the paper, JA, KB, MN, NF and AA conducted the metabolite quantification. MP, GK and NF performed the bioinformatics and statistical analysis. The recruitment and management of the study was led by HV, TK, WR and MN. CW and MP designed the study. All authors read and revised the manuscript and gave final approval. MP serves as guarantor of this work and is responsible for the integrity of the data and its analysis.

Disclosure

The authors declare no conflict of interest.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. WHO, *Global Report on Diabetes*. 2016, World Health Organization: Geneva.
2. WHO, *Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO consultation.* 2006, World Health Organization: Geneva.
3. *Standards of medical care in diabetes--2012.* Diabetes Care, 2012. **35 Suppl 1**: p. S11-63.
4. Cobb, J., et al., *A novel test for IGT utilizing metabolite markers of glucose tolerance*. J Diabetes Sci Technol, 2015. **9**(1): p. 69-76.
5. Rich, P.A., et al., *Using a quantitative measure of diabetes risk in clinical practice to target and maximize diabetes prevention interventions.* Clinical Diabetes, 2013. **31**(2): p. 82-89.
6. Peddinti, G., et al., *Early metabolic markers identify potential targets for the prevention of type 2 diabetes*. Diabetologia, 2017. **60**(9): p. 1740-1750.
7. Wang, T.J., et al., *Metabolite profiles and the risk of developing diabetes.* Nat Med, 2011. **17**(4): p. 448-53.
8. Wang-Sattler, R., et al., *Novel biomarkers for pre-diabetes identified by metabolomics.* Mol Syst Biol, 2012. **8**: p. 615.
9. Yousri, N.A., et al., *A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control.* Diabetologia, 2015. **58**(8): p. 1855-67.
10. Merino, J., et al., *Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose*. Diabetologia, 2018. **61**(6): p. 1315-1324.
11. Rebholz, C.M., et al., *Serum metabolomic profile of incident diabetes.* Diabetologia, 2018. **61**(5): p. 1046-1054.
12. Palmer, N.D., et al., *Metabolomics Identifies Distinctive Metabolite Signatures for Measures of Glucose Homeostasis: The Insulin Resistance Atherosclerosis Family Study (IRAS-FS)*. J Clin Endocrinol Metab, 2018. **103**(5): p. 1877-1888.
13. Wurtz, P., et al., *Circulating metabolite predictors of glycemia in middle-aged men and women*. Diabetes Care, 2012. **35**(8): p. 1749-56.
14. Guasch-Ferre, M., et al., *Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis*. Diabetes Care, 2016. **39**(5): p. 833-46.
15. Lanza, I.R., et al., *Quantitative metabolomics by H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes.* PLoS One, 2010. **5**(5): p. e10538.
16. Menni, C., et al., *Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach*. Diabetes, 2013. **62**(12): p. 4270-6.
17. Gall, W.E., et al., *alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population.* PLoS One, 2010. **5**(5): p. e10883.
18. Ahlqvist, E., et al., *Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables.* Lancet Diabetes Endocrinol, 2018. **6**(5): p. 361-369.
19. Li, L., et al., *Identification of type 2 diabetes subgroups through topological analysis of patient similarity*. Sci Transl Med, 2015. **7**(311): p. 311ra174.
20. Udler, M.S., et al., *Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis.* PLoS Med, 2018. **15**(9): p. e1002654.
21. McCarthy, M.I., *Painting a new picture of personalised medicine for diabetes*. Diabetologia, 2017. **60**(5): p. 793-799.
22. Volzke, H., et al., *Cohort profile: the study of health in Pomerania.* Int J Epidemiol, 2011. **40**(2): p. 294-307.

23. Matthews, D.R., et al., *Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man*. Diabetologia, 1985. **28**(7): p. 412-9.
24. Pietzner, M., et al., *Hepatic Steatosis Is Associated With Adverse Molecular Signatures in Subjects Without Diabetes*. J Clin Endocrinol Metab, 2018. **103**(10): p. 3856-3868.
25. Charrad, M., et al., *Package 'nbclust'*. Journal of statistical software, 2014. **61**: p. 1-36.
26. Krumsiek, J., et al., *Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data*. BMC Syst Biol, 2011. **5**: p. 21.
27. Dousset, B., et al., *Purification from human plasma of a hexapeptide that potentiates the sulfation and mitogenic activities of insulin-like growth factors*. Biochem Biophys Res Commun, 1998. **247**(3): p. 587-91.
28. Guertin, K.A., et al., *Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cance*r. Am J Clin Nutr, 2015. **101**(5): p. 1000-11.
29. Allalou, A., et al., *A Predictive Metabolic Signature for the Transition From Gestational Diabetes Mellitus to Type 2 Diabetes*. Diabetes, 2016. **65**(9): p. 2529-39.
30. Cobb, J., et al., *alpha-Hydroxybutyric Acid Is a Selective Metabolite Biomarker of Impaired Glucose Tolerance*. Diabetes Care, 2016. **39**(6): p. 988-95.
31. Lynch, C.J. and S.H. Adams, *Branched-chain amino acids in metabolic signalling and insulin resistance.* Nat Rev Endocrinol, 2014. **10**(12): p. 723-36.
32. Pedersen, H.K., et al., *Human gut microbes impact host serum metabolome and insulin sensitivity*. Nature, 2016. **535**(7612): p. 376-81.
33. Dashty, M., *A quick look at biochemistry: carbohydrate metabolism.* Clin Biochem, 2013. **46**(15): p. 1339-52.
34. Park, J.E., et al., *Metabolite changes in risk of type 2 diabetes mellitus in cohort studies: A systematic review and meta-analysis.* Diabetes Res Clin Pract, 2018. **140**: p. 216-227.
35. Wang, Q., et al., *Genetic Support for a Causal Role of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation*. Diabetes Care, 2017. **40**(12): p. 1779-1786.
36. Lotta, L.A., et al., *Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis*. PLoS Med, 2016. **13**(11): p. e1002179.
37. Ferrannini, E., et al., *Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance*. Diabetes, 2013. **62**(5): p. 1730-7.
38. Floegel, A., et al., *Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach*. Diabetes, 2013. **62**(2): p. 639-48.
39. Semba, R.D., et al., *Altered Plasma Amino Acids and Lipids Associated With Abnormal Glucose Metabolism and Insulin Resistance in Older Adults*. J Clin Endocrinol Metab, 2018. **103**(9): p. 3331-3339.
40. Cole, L.K., J.E. Vance, and D.E. Vance, *Phosphatidylcholine biosynthesis and lipoprotein metabolism.* Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 2012. **1821**(5): p. 754-761.
41. Ginsberg, H.N. and L.-S. Huang, *The insulin resistance syndrome: impact on lipoprotein metabolism and atherothrombosis.* Journal of cardiovascular risk, 2000. **7**(5): p. 325-331.

42. Pietzsch, J. and K. Fuecker, *Increased cholesteryl ester transfer protein activity in impaired glucose tolerance: relationship to high density lipoprotein metabolism*. Croatian medical journal, 2002. **44**(2): p. 171-177.
43. Choi, S.H. and H.N. Ginsberg, *Increased very low density lipoprotein (VLDL) secretion, hepatic steatosis, and insulin resistance*. Trends in Endocrinology & Metabolism, 2011. **22**(9): p. 353-363.
44. Dullaart, R.P., et al., *Increased large VLDL particles confer elevated cholesteryl ester transfer in diabetes*. Eur J Clin Invest, 2015. **45**(1): p. 36-44.
45. Forouhi, N.G., et al., *Association of Plasma Phospholipid n-3 and n-6 Polyunsaturated Fatty Acids with Type 2 Diabetes: The EPIC-InterAct Case-Cohort Study*. PLoS Med, 2016. **13**(7): p. e1002094.
46. Abbasi, A., et al., *Bilirubin as a potential causal factor in type 2 diabetes risk: a Mendelian randomization study*. Diabetes, 2015. **64**(4): p. 1459-69.
47. Cheriyath, P., et al., *High Total Bilirubin as a Protective Factor for Diabetes Mellitus: An Analysis of NHANES Data From 1999 - 2006*. J Clin Med Res, 2010. **2**(5): p. 201-6.
48. Ndisang, J.F., *Role of heme oxygenase in inflammation, insulin-signalling, diabetes and obesity*. Mediators Inflamm, 2010. **2010**: p. 359732.
49. Wlazlo, N., et al., *Complement factor 3 is associated with insulin resistance and with incident type 2 diabetes over a 7-year follow-up period: the CODAM Study*. Diabetes Care, 2014. **37**(7): p. 1900-9.
50. Holland, W.L., et al., *Lipid Mediators of Insulin Resistance*. Nutrition Reviews, 2007. **65**(suppl_1): p. S39-S46.
51. Senn, J.J., *Toll-like receptor-2 is essential for the development of palmitate-induced insulin resistance in myotubes*. J Biol Chem, 2006. **281**(37): p. 26865-75.
52. Pradhan, A.D., et al., *C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus*. Jama, 2001. **286**(3): p. 327-34.
53. Samuel, V.T. and G.I. Shulman, *Mechanisms for insulin resistance: common threads and missing links*. Cell, 2012. **148**(5): p. 852-71.
54. Schooneman, M.G., et al., *Acylcarnitines: reflecting or inflicting insulin resistance?* Diabetes, 2013. **62**(1): p. 1-8.
55. Qiu, G., et al., *Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults*. Int J Epidemiol, 2016. **45**(5): p. 1507-1516.
56. Nowak, C., et al., *Effect of Insulin Resistance on Monounsaturated Fatty Acid Levels: A Multi-cohort Non-targeted Metabolomics and Mendelian Randomization Study*. PLoS Genet, 2016. **12**(10): p. e1006379.
57. Pietzner, M., et al., *Comprehensive metabolic profiling of chronic low-grade inflammation among generally healthy individuals*. BMC Med, 2017. **15**(1): p. 210.

Fig. 1 Flow chart of sample compilation and the statistical analysis. Plasma metabolome data based on MS and ^{1}H-NMR were available for 1000 subjects. Exclusion criteria applied and the number of excluded subjects are indicated. A total sample of 827 subjects was included in the analyses using linear regression models to assess associations of plasma metabolites with 2-hour measures of plasma glucose and insulin. IGT cases (n=159) were selected and subjected to unsupervised clustering using a k-means or hierarchical clustering approach (HCA). A signature predictive for impaired glucose tolerance using least absolute shrinkage and selection operator (LASSO) for variable selection was compiled. Three types of variable set ups were used to perform this classification (only clinical variables, only

metabolites significantly associated with IGT, a combination of both). Three sparse logistic regression models were built to predict impaired glucose tolerance.

Fig. 2 Standardized β-estimates with 95%-confidence intervals from linear regression analysis predicting either 2-hour glucose (squares) or insulin (circles) concentrations with plasma metabolites. All analyses were adjusted for baseline concentrations, age, sex, smoking, physical activity, waist circumference, estimated glomerular filtration, serum alanine aminotransferase activities and high-sensitivity C-reactive protein. Significant associations (controlling the false discovery rate at 5%) are highlighted by darker colours. More detailed information on the metabolites could be found in table 2.

Fig. 3 A) Comparison of effect estimates from linear regression analysis (2-hour glucose concentrations as exposure) with odds ratios (OR) form logistic regression analysis for impaired glocuse tolearnce (IGT, 2-hour glucose > 7.8 mmol/l). **B)** Standardized β-estimates with 95%-confidence intervals from linear regression analysis predicting either 2-hour glucose (squares) or insulin (circles) concentrations with lipoprotein subfraction measures. All analyses were adjusted for baseline concentrations, age, sex, smoking, physical activity, waist circumference, estimated glomerular filtration, serum alanine aminotransferase activities and high-sensitivity C-reactive protein. Significant associations (controlling the false discovery rate at 5%) are highlighted by darker colours. VLDL = very low-density lipoprotein; LDL = low-density lipoprotein; ApoB = apolipoprotein B; HDL = high-density lipoprotein

Fig. 4 *left* Heatmap of standardized metabolite levels (x-axis) sorted based on hierarchical clustering. Columns next to the heatmap indicate correspondence with k-means clustering and indication of subjects with 2-hour glucose concentrations > 11.1 mmol/l reflecting type 2 diabetes mellitus. ***right*** Boxplots of parameters for glucose homeostasis according to cluster membership (k-means). Stars indicate significant differences (Welchs t-test).

Fig. 5 A) Receiver operating characteristics curves and area under the curve (AUC) with 95%-confidence interval (CI) for the three different models to predict impaired glucose tolerance: Clinical variables – dark grey; Metabolites – light grey; Combination of both – black. **B)** Subnetwork of the derived metabolic network centred on X – 11727. Nodes indicate metabolites and edges significant partial correlations. Line width corresponds to strength of partial correlations.

Table 1 Characteristics of the study population by glucose tolerance

Characteristic	Normoglycaemic (n = 668)	Impaired Glucose Tolerance (n = 159)	p*
Age (years)	49 (40; 58)	60 (50; 68)	<0.01
Females (%)	57.3	52.8	0.35
Smoking (%)			<0.01
never smoker	41.4	52.2	
former smoker	36.1	39.6	
current smoker	22.5	8.2	
Physically active (%)	73.3%	75.5%	0.66
Waist circumference (cm)	86 (77; 95)	95 (85; 102)	<0.01
Glucose 0 hour (mmol/L)	5.3 (5.0; 5.6)	5.8 (5.3; 6.3)	<0.01
Glucose 2 hour (mmol/L)	5.8 (5.1; 6.5)	8.9 (8.2; 10.5)	<0.01
Insulin 0 hour (µU/mL)	8.3 (6.0; 11.9)	13.5 (9.1; 18.1)	<0.01
Insulin 2 hour (µU/mL)	45.0 (30.0; 65.6)	134.0 (76.4; 184.3)	<0.01
HOMA-IR	1.97 (1.40; 2.90)	3.4 (2.2; 4.9)	<0.01
HbA_{1c} (%)	5.1 (4.8; 5.4)	5.4 (5.1; 5.7)	<0.01
Triglycerides (mmol/L)	1.16 (0.84; 1.61)	1.46 (1.09; 1.96)	<0.01
LDL-cholesterol (mmol/L)	3.36 (2.76; 3.98)	3.47 (2.90; 4.00)	0.27
Total cholesterol (mmol/L)	5.4 (4.8; 6.2)	5.5 (4.9; 6.1)	0.17

hsCRP, mg/L	1.08 (0.57; 2.12)	1.66 (0.94; 3.90)	<0.01
ALT (µkatal/L)	0.36 (0.27; 0.50)	0.42 (0.31; 0.60)	<0.01
eGFR (mL/min/1.73m²)	114 (106; 123)	108 (99; 117)	<0.01

HOMA-IR = homeostatic model of insulin resistance; hsCRP = high-sensitivity C-reactive protein; LDL = low-density lipoprotein ALT = alanine transaminase; eGFR = estimated glomerular filtration rate; Continuous data are expressed as median (25^{th} percentile; 75^{th} percentile); nominal data are given as percentages. *χ2-test (nominal data) or Mann-Whitney-U test (interval data) were performed

Table 2 Additional information for metabolites significantly associated with OGTT-measures.

Metabolite	HMDB ID	P.	D.	Metabolite	HMDB ID	P.	D.
Alanine	HMDB00161	B	+	lysoPC a C18:1	HMDB02815	B	-
Asparagine	HMDB00168	B	-	lysoPC a C18:2	HMDB10386	B	-
Betaine	HMDB00043	M	-	5-dodecenoate (12:1n7)	HMDB00529	M	+
Glycine	HMDB00123	B	-	Caprate (10:0)	HMDB00511	M	+
3-hydroxyisobutyrate	HMDB00336	M	+	Caprylate (8:0)	HMDB00482	M	+
3-methyl-2-oxobutyrate	HMDB00019	M	+	Laurate (12:0)	HMDB00638	M	+
3-methyl-2-oxovalerate	HMDB03736	M	+	PC ae C32:1	HMDB13404	B	-
4-methyl-2-oxopentanoate	HMDB00695	M	+	PC ae C32:2	HMDB13411	B	-
Alpha-hydroxyisovalerate	HMDB00407	M	+	PC ae C34:3	HMDB11211	B	-
Beta-hydroxyisovalerate	HMDB00754	M	+	PC ae C36:2	HMDB11243	B	-
Isobutyrylcarnitine	HMDB00736	M	-	PC ae C40:6	HMDB13422	B	-
Indoleacetate	HMDB00197	M	-	PC ae C42:5	HMDB13451	B	-
Kynurenine	HMDB00684	B	-	PC ae C44:5	HMDB13456	B	-
Citrulline	HMDB00904	B	-	PC aa C30:2	HMDB07999; HMDB07903	B	-
Ornithine	HMDB00214	B	-	PC aa C36:2	HMDB00593	B	-
Lactate	HMDB00190	M	+	PC aa C36:4	HMDB07982	B	+
Pyruvate	HMDB00243	M	+	PC aa C38:0	HMDB07893	B	-
Bilirubin (E,E)*	HMDB00054	M	+	Adrenate (22:4n6)	HMDB02226	M	+
L-urobilin	HMDB04159	M	+	Arachidonate (20:4n6)	HMDB01043	M	+
Trigonelline (N'-methylnicotinate)	HMDB00875	M	-	Dihomo-linoleate (20:2n6)	HMDB05060	M	+
Phosphate	HMDB01429	M	-	Dihomo-linolenate (20:3n3 or n6)	HMDB02925	M	+
Citrate	HMDB00094	M	+	Docosadienoate (22:2n6)	HMDB61714	M	+
Carnitine	HMDB00062	B	-	Docosahexaenoate (DHA; 22:6n3)	HMDB02183	M	+
Acetylcarnitine	HMDB00201	B	+	Docosapentaenoate (n3 DPA; 22:5n3)	HMDB01976	M	+
Cis-4-decenoyl carnitine	57357170*	M	+	Docosapentaenoate (n6 DPA; 22:5n6)	HMDB13123	M	+
Dodecanoylcarnitine	HMDB02250	B	+	Eicosapentaenoate (EPA; 20:5n3)	HMDB01999	M	+
Hexanoylcarnitine	HMDB00705	M	+	Linoleate (18:2n6)	HMDB00673	M	+
Hydroxybutyrylcarnitine*	HMDB13127	M	+	Linolenate [alpha or gamma; (18:3n3 or 6)]	HMDB01388	M	+
Octanoylcarnitine	HMDB00791	M	+	SM (OH) C16:1	HMDB13463	B	-
15-methylpalmitate	HMDB61709	M	+	SM (OH) C22:2	HMDB13467	B	-
17-methylstearate	HMDB61710	M	+	SM (OH) C24:1	HMDB12107	B	-
13-HODE + 9-HODE	HMDB04667; HMDB10223	M	+	SM C16:0	HMDB10169	B	-
2-hydroxypalmitate	HMDB31057	M	+	SM C24:1	HMDB12107	B	-
3-hydroxydecanoate	HMDB02203	M	+	Phenylalanyltryptophan	HMDB29006	M	+
3-hydroxyoctanoate	HMDB01954	M	+	HWESASLLR	Peptide fragments	M	+
3-hydroxybutyrate (BHBA)	HMDB00357	M	+	HWESASXX*		M	+
10-heptadecenoate (17:1n7)	HMDB60038	M	+	4-vinylphenol sulfate	HMDB04072	M	-
10-nonadecenoate (19:1n9)	HMDB13622	M	+	Catechol sulfate	HMDB59724	M	-
Arachidate (20:0)	HMDB02212	M	+	Hippurate	HMDB00714	M	-
Eicosenoate (20:1n9 or 11)	HMDB02231	M	+	X - 11381		M	-
Erucate (22:1n9)	HMDB02068	M	+	X - 11440		M	+
Margarate (17:0)	HMDB02259	M	+	X - 11793		M	+
Myristate (14:0)	HMDB00806	M	+	X - 11977		M	+
Myristoleate (14:1n5)	HMDB02000	M	+	X - 12216		M	-
Nonadecanoate (19:0)	HMDB00772	M	+	X - 12742		M	+
Oleate (18:1n9)	HMDB00207	M	+	X - 12816		M	-
Palmitate (16:0)	HMDB00220	M	+	X - 12855		M	+
Palmitoleate (16:1n7)	HMDB03229	M	+	X - 17357		M	+
Stearate (18:0)	HMDB00827	M	+	X - 17359		M	+
1-linoleoylglycerophosphoethanolamine*	HMDB11507	M	-				

HMDB ID = identifier for the human metabolome data base (www.hmdb.ca); P. = platform metabolie was measured (M = Metabolon, B = Biocrates); D. = direction of association with 2-hour glucose levels (+ = positively; - = inversely); *PubChem identifier, since not listed in HMDB

Table 3 Summary on predictors selected for the final classification scheme discriminating impaired glucose tolerance.

Clinical Traits			Metabolites			Combined		
Variable	**Score**	**OR (95%-CI)**	**Variable**	**Score**	**OR (95%-CI)**	**Variable**	**Score**	**OR (95%-CI)**
Fasting glucose	0.77	2.49 (2.06;3.03)	Alanine	0.77	1.63 (1.36;1.97)	Fasting glucose	0.80	2.49 (2.06;3.03)
hsCRP	0.75	1.56 (1.31;1.85)	lysoPC a C18:2	0.77	0.55 (0.45;0.65)	Age	0.75	2.07 (1.70;2.53)
Age	0.62	2.07 (1.70;2.53)	Hexoses	0.69	2.12 (1.76;2.58)	lysoPC a C18:2	0.72	0.55 (0.45;0.65)
HOMA-IR	0.49	2.20 (1.84;2.67)	X - 11727	0.51	1.97 (1.63;2.40)	Alanine	0.67	1.63 (1.36;1.97)
Current smoking	0.33	0.63 (0.48;0.78)	2-oleoylglycerophosphocholine	0.33	0.79 (0.66;0.94)	3-methyl-2-oxobutyrate	0.43	1.89 (1.57;2.30)
						HOMA-IR	0.40	2.20 (1.84;2.67)
						trigonelline	0.37	0.75 (0.63;0.90)
						2-oleoylglycerophosphocholine	0.32	0.79 (0.66;0.94)

OR (95%-CI) = crude odds ratio per standard deviation increase for impaired glucose tolerance with 95%-confidence interval derived from logistic regression models; Score = defined as average area under the curve in the final classification loop in case the variable was included in the LASSO model (see Methods); HOMA-IR = homeostatic model assessment of insulin resistance; hsCRP = high-sensitivity C-reactive protein; lysoPC = Lysophosphatidylcholine